*S,N 104*
*5 114 04*

# Watermark Enabled Video Objects

*This application is continuation of 09/597,209 filed 6/20/2000, now pat. 6,411,725 which is a CIP of 09/408,902 filed 9/29/99, Now pat 6,408,331 which is a DIV of 09/130,624 filed 8/6/98 Now pat. 6,324,573, which is CON of 08/508,083 filed 7/27/95 now pat. 5,841,978.*

## Related Application Data

The subject matter of the present application is related to that disclosed in US
Patent 5,862,260, and in co-pending U.S. Patent Applications:

5          09/503,881, filed February 14, 2000;

           60/082,228, filed April 16, 1998;

           09/292,569, filed April 15, 1999;

           60/134,782, filed May 19, 1999;

           09/343,104, filed June 29, 1999;

10         60/141,763, filed June 30, 1999;

           09/562,517, filed May 1, 2000;

           09/531,076, filed March 18, 2000and

           09/571,422, filed May 15, 2000;

           which are hereby incorporated by reference.

15

## Technical Field

The invention relates to multimedia signal processing, and in particular relates to
encoding information into and decoding information from video objects.

## Background and Summary

20         "Steganography" refers to methods of hiding auxiliary information in other
information.  Audio and video watermarking are examples of steganography.  Digital
watermarking is a process for modifying media content to embed a machine-readable
code into the data content.  A media signal, such as an image or audio signal, is modified
such that the embedded code is imperceptible or nearly imperceptible to the user, yet may

25  be detected through an automated detection process.  Most commonly, digital
watermarking is applied to media such as images, audio signals, and video signals.

However, it may also be applied to other types of data, including documents (e.g., through line, word or character shifting), software, multi-dimensional graphics models, and surface textures of objects.

Digital watermarking systems have two primary components: an embedding component that embeds the watermark in the media content, and a reading component that detects and reads the embedded watermark. The embedding component embeds a watermark by altering data samples of the media content. The reading component analyzes content to detect whether a watermark is present. In applications where the watermark encodes information, the reader extracts this information from the detected watermark.

The invention provides methods and systems for associating video objects in a video sequence with object specific actions or information using auxiliary information embedded in video frames or audio tracks. A video object refers to a spatial and temporal portion of a video signal that depicts a recognizable object, such as a character, prop, graphic, etc. Each frame of a video signal may have one or more video objects. The auxiliary information is embedded in video or audio signals using "steganographic" methods, such as digital watermarks. By encoding object specific information into video or an accompanying audio track, the watermarks transform video objects into "watermark enabled" video objects that provide information, actions or links to additional information or actions during playback of a video or audio-visual program. A similar concept may be applied to audio objects, i.e. portions of audio that are attributable to a particular speaker, character, instrument, artist, etc.

One aspect of the invention is a method for encoding substantially imperceptible auxiliary information about a video object into a video signal that includes at least one video object. The method steganographically encodes object specific information about the video object into the video signal. Some examples of this information include identifiers and screen locations of corresponding video objects. The method associates the object specific information with an action. This action is performed automatically or

in response to user selection of the video object through a user interface while the video signal is playing.

Another aspect of the invention is a method for encoding substantially imperceptible auxiliary information into physical objects so that the information survives

5    the video capture process and links the video to an action. This method steganographically encodes auxiliary information in a physical object in a manner that enables the auxiliary information to be decoded from a video signal captured of the physical object. One example is to place a watermarked image on the surface of the object. The method associates the auxiliary information with an action so that the video

10   signal captured of the physical object is linked to the action. One example of an action is retrieving and displaying information about the object. For example, the watermark may act as a dynamic link to a web site that provides information about the object.

Another aspect of the invention is a method for using a watermark that has been encoded into a video signal or in an audio track accompanying the video signal. The

15   watermark conveys information about a video object in the video signal. The method decodes the information from the watermark, receives a user selection of the video object, and executes an action associated with the information about the video object. One example of an action is to retrieve a web site associated with the video object via the watermark. The watermark may include a direct (e.g., URL or network address) or

20   indirect link (e.g., object identifier) to the web site. In the latter case, the object identifier may be used to look up a corresponding action, such as issuing a request to a web server at a desired URL. Object information returned to the user (e.g., web page) may be rendered and superimposed on the same display as the one displaying the video signal, or a separate user interface.

25   Another aspect of the invention is a system for creating watermark enabled video objects. The system includes an encoder for encoding a watermark in a video sequence or accompanying audio track corresponding to a video object or objects in the video sequence. It also includes a database system for associating the watermark with an action

or information such that the watermark operable to link the video object or objects to a related action or information during playback of the video sequence.

Another aspect of the invention is a system for processing a watermark enabled video object in a video signal. The system comprises a watermark decoder and rendering

5      system. The watermark decoder decodes a watermark carrying object specific information from the video signal and linking object specific information to an action or information. The rendering system renders the action or information.

Another aspect of the invention is a method for encoding substantially imperceptible auxiliary information into an audio track of a video signal including at least

10     one video object. This method steganographically encodes object specific information about the video object into the audio track. It also associates the object specific information with an action, where the action is performed in response to user selection of the video object through a user interface while the video signal is playing. Alternatively, the action can be performed automatically as the video is played.

15     Further features will become apparent with reference to the following detailed description and accompanying drawings.

## Brief Description of the Drawings

Fig. 1A is a flow diagram depicting a process for encoding and decoding

20     watermarks in content to convey auxiliary information 100 about video objects in the content.

Fig. 1B illustrates a framework outlining several alternative implementations of linking video objects with actions or information.

Fig. 2 is a flow diagram depicting a video creation process in which physical

25     objects are pre-watermarked in a manner that survives video capture and transmission.

Fig. 3 is a flow diagram of a video creation process that composites watermarked video objects with a video stream to create a watermarked video sequence.

Fig. 4 illustrates an embedding process for encoding auxiliary information about video objects in a video stream.

Fig. 5 is a diagram depicting yet another process for encoding auxiliary information about video objects in a video stream.

Fig. 6 depicts an example watermark encoding process.

Fig. 7 is a diagram depicting decoding processes for extracting watermark

5    information from video content and using it to retrieve and render external information or actions.

Fig. 8 illustrates an example configuration of a decoding process for linking video objects to auxiliary information or actions.

Fig. 9 illustrates another example configuration of a decoding process for linking

10   video objects to auxiliary information or actions.

## Detailed Description

The following sections detail ways to encode and decode information, actions and links into video objects in a video sequence. A video object refers to a video signal depicting an object of a scene in a video sequence.

15   To a viewer, the video object is recognizable and distinguishable from other imagery in the scene. The video object exists in a video sequence for some duration, such as a contiguous set of video frames. A single image instance in a frame corresponding to the object is a video object layer. The video object may comprise a sequence of natural images that occupy a portion of each frame in a video sequence, such

20   as a nearly static talking head or a moving athlete. Alternatively, the video object may be a computer generated rendering of a graphical object that is layered with other renderings or natural images to form each frame in a video sequence. In some cases, the video object may encompass an entire frame.

In the systems described below, watermarks are encoded and decoded from video

25   or audio tracks for the purpose of conveying information related to the video objects. A watermark encoding process embeds a watermark into an audio or video signal, or in some cases, the physical object that later becomes a video object through video capture. At playback, a decoding process extracts the watermark.

Fig. 1A is a flow diagram depicting a process for encoding and decoding watermarks in content to convey auxiliary information 100 about video objects in the content. An embedding process 102 encodes the auxiliary information into a watermark embedded in the video content. A transmitter 104 then distributes the content to viewers, via broadcast, electronic file download over a network, streaming delivery over a network, etc. A receiver 106 captures the video content and places it in a format from which a watermark decoder 108 extracts the auxiliary information. A display 110 displays the video to a viewer. As the video is being displayed, a user interface 114 executes and provides visual, audio, or audio-visual information to the user indicating that the video is embedded with auxiliary information or actions. This user interface may be implemented by superimposing graphical information over the video on the display 110. Alternatively, the decoder can pass auxiliary object information to a separate device, which in turn, executes a user interface. In either case, the user interface receives input from the user, selecting a video object. In response, it performs an action associated with the selected object using the auxiliary object information decoded from the watermark.

The watermark may carry information or programmatic action. It may also link to external information or an action, such as retrieval and output of information stored elsewhere in a database, website, etc. Watermark linking enables the action associated with the watermark to be dynamic. In particular, the link embedded in the content may remain the same, but the action or information it corresponds to may be changed.

Watermark linking of video objects allows a video object in a video frame to trigger retrieval of information or other action in response to selection by a user. Watermark embedding may be performed at numerous and varied points of the video generation process. For 3D animation, the watermark can be embedded immediately into a video object layer after a graphical model is rendered to the video object layer, allowing a robust and persistent watermark to travel from the encoding device or computer to any form of playback of video containing the video object.

For special effects, an actor filmed against a green screen can be embedded directly after the film is transferred to digital format for effects generation, preventing the need to later extract the actor from the background to embed only his image. For network or cable broadcast news, the ubiquitous pop-up screen that appears next to the

5    news anchor's head can be embedded before the newscast allowing the viewer to click on that image to take them to extra information from a website.

Watermarks may be embedded in broadcast video objects in real time. An example is watermarking NBA basketball players as a game is broadcast allowing the view to click on players and receive more information about them.

10   Wherever the video is distributed, a decoding process may be inserted to decode information about the video object from a watermark embedded in the video signal. This information may then be used to trigger an action, such as fetching graphics and displaying it to the user. For example, the watermark information may be forwarded to a database, which associates an action with the watermark information. One form of such a

15   database is detailed in co-pending application 09/571,422, which is hereby incorporated by reference. This database looks up an action associated with watermark information extracted from content. One action is to issue a query to a web server, which in turn, returns a web page to the user via the Internet, or some other communication link or network.

20   Fig. 1B illustrates a system architecture outlining several alternative implementations of linking video objects with actions or information. This diagram divides the system into a creation side, where content is created and encoded, and an end user side, where content and watermark enabled information or actions are rendered. On the creation side, the diagram shows examples of three watermark types and two

25   watermark protocols. In type one, the watermark is embedded in a physical object before it is recorded in a video signal. In type two, the watermark is encoded in a video object after it is recorded but before it is broadcast, possibly during a video editing process. For example, this type of watermark may be encoded in a video object of an actor captured in front of a greenscreen as he moves through a scene. In type three, the watermark is

added as the video is being captured for a live event, such as watermarking a video object depicting the jersey of a basketball player as a video stream is being captured of a game.

In the first protocol, the watermark is encoded in the video frame area of the desired object, such as where the jersey of the basketball player appears on the video

5  display screen.  In the second protocol, the watermark is encoded throughout a video frame or corresponding segment of an audio track, and includes information about the object and its location.  For example, during the basketball game, the watermark is embedded in the audio track and includes location, size and identification for player 1, then player 2, then player 3, and back to player 1 if he is still in the scene or onto player 2

10  or 3, etc.

On the end user side, there are two places for network connectivity, rendering of linked information, and user interaction.  Internet connectivity can be included in the video display device or associated set-top box or in a portable display device, such as a personal laptop.  The rendering of the linked information can occur on the video display,

15  possibly using picture-in-picture technology so others can still see the original video, or in the portable display device, such as a laptop since Internet browsing can be a personal experience.  User interaction with the system, such as selecting the object to find linked information can happen with the video display, such as pointing with a remote, or with a portable display device, such as using a mouse on a laptop.  Specific implementations can

20  include a variety of combination of these components.

### Embedding Processes

The embedding process encodes one or more watermarks into frames of a video sequence, or in some cases, an audio track that accompanies the video sequence.  These watermarks carry information about at least one video object in the sequence, and also

25  create an association between a video object and an action or external information.  The association may be formed using a variety of methods.

One method is to encode an object identifier in a watermark.  On the decoding side, this identifier is used as a key or index to an action or information about a video object.  The identifier may be a direct link to information or actions  (e.g., an address of

the information or action), or be linked to the information or actions via a server database.

Another method is to encode one object identifier per frame, either in the frame or corresponding audio track segment. Then, the system sends a screen location selected by a user and the identifier to the server. The object identifier plays a similar role as the previous method, namely, it identifies the object. The location information may be used along with the object identifier to form an index into a database to look up a database entry corresponding to a video object.

Alternatively, the watermark may contain several identifiers and corresponding locations defining the screen location of a related video object. The screen location selected by the user determines which identifier is sent to the server for linked information or actions. In other words, a process at the end-user side maps the location of the user selection to an identifier based on the locations encoded along with the identifiers in the content. For example, a segment of the audio track that is intended to be played with a corresponding video frame or frame sequence may include a watermark or watermarks that carry one or more pairs of identifier and locations. These watermarks may be repeated in audio segments synchronized with video frames that include corresponding linked video objects. Then, in the decoding process, the identifier closest to the location of the user interaction is used. A modification includes providing bounding locations in the watermark and determining whether the user's selection is within this area, as opposed to using the closest watermark location to the user's selection.

Other context information available at decoding time may be used to create an association between a video object in a frame and a corresponding action or information in a database. For example, the frame number, screen coordinates of a user selection, time or date may be used in conjunction with information extracted from the watermark to look up a database entry corresponding to a video object in a video sequence.

The manner in which the embedded data is used to create an association between video objects and related information or actions impacts how that data is embedded into

each frame. For example, if the watermark includes location information, an object identifier can be embedded throughout the frame in which the corresponding object resides, rather than being located in a portion of the frame that the object overlaps. If the frame includes two or more linked video objects, the watermark conveys an object

5   identifier and location for each of the video objects.

Additional decoding side issues impacting the encoding process include: 1) enabling the user to select video objects during playback; and 2) mapping a user's input selecting a video object to the selected video object. The user can select a video object in various ways. For example, gestural input devices, such as a mouse, touch screen, etc.

10   enable the user to select a video object by selecting a screen location occupied by that object. The selected location can then be mapped to information extracted from a watermark, such as an object identifier. The object identifier of a video object that overlaps the selected location can be looked up based on location codes embedded in the watermark or by looking up the object identifier extracted from a watermark in a video

15   object layer at the selected location.

If a user interface on the decoding side provides additional information about watermarked video objects, like graphical icons, menus, etc., then the user can select a video object by selecting a graphic, menu item, or some other user interface element associated with that object. There are many ways to select graphics or menu items,

20   including gestural input devices, keyboards, speech recognition, etc. This approach creates an additional requirement that the decoding side extract watermark information and use it to construct a graphical icon or menu option to the user. The decoding process may derive the information needed for this user interface from the video content, from a watermark in the content, or from out-of-band auxiliary data. In the latter two

25   cases, the embedding process encodes information into the content necessary to generate the user interface on the decoding side.

An example will help illustrate an encoding process to facilitate user selection of video objects on the decoding side. Consider an example where a watermark encoder encodes a short title (or number) and location of marked video objects into the video

stream containing these objects. The decoding process can extract the title and location information, and display titles at the locations of the corresponding video objects. To make the display less obtrusive to the playback of the video, the display of this auxiliary information can be implemented using small icons or numbers superimposed on the video during playback, or it can be transmitted to a separate device from the device displaying the video. For example, the video receiver can decode the information from the video stream and send it via wireless transmission to an individual user's hand held computer, which in turn, displays the information and receives the user's selection.

There a number of different embedding scenarios for encoding information into a video stream to link video objects with information or actions. Figs. 2-5 illustrate some examples. In Fig. 2, physical objects 200 are pre-watermarked in a manner that survives the video capture process 202. For an example of a watermarking process that survives digital to analog conversion (e.g., printing a digital image on a physical object), and then analog to digital conversion (e.g., capture via a video camera), see US Patent 5,862,260, and in co-pending patent application 09/503,881, filed February 14, 2000. These approaches are particularly conducive but not limited to applications where the objects are largely flat and stationary, such as billboards, signs, etc. The video capture process records the image on the surface of these objects, which is encoded with a watermark. The resulting video is then transmitted or broadcast 204.

In the process of Fig. 3, a video creation process composites watermarked video objects 300 with a video stream 302 to create a watermarked video sequence. The watermark may be encoded into video object layers. Examples of watermark encoding and decoding technology are described in US Patent 5,862,260, and in co-pending applications 09/503,881, filed February 14, 2000, and WO 99/10837.

A compositing operation 304 overlays each of the video objects onto the video stream in depth order. To facilitate automated compositing of the video object layers, each of the objects has depth and transparency information (e.g., sometimes referred to as translucency, opacity or alpha). The depth information indicates the relative depth ordering of the layers from a viewpoint of the scene (e.g., the camera position) to the

background.  The transparency indicates the extent to which pixel elements in a video object layer allow a layer with greater depth to be visible.  The video generated from the compositing operation sequence is broadcast or transmitted to viewers.

The video objects may be encoded with watermarks as part of a compression

5    process.  For example, the MPEG 4 video coding standard specifies a video compression codec in which video object layers are compressed independently.  In this case, the video object layers need not be composited before they are transmitted to a viewer.  At the time of viewing, an MPEG 4 decoder  decompresses the video object layers and composites them to reconstruct the video sequence.

10    The watermark may be encoded into compressed video object layers by modulating DCT coefficients of intra or interframe macroblocks.  This watermark can be extracted from the DCT coefficients before the video objects are fully decompressed and composited.

Fig. 4 illustrates another embedding process for encoding auxiliary information

15    about video objects in a video stream 400.  In this embedding process, a user designates a video object and the auxiliary information to be encoded in the video object via a video editing tool 402.  A watermark encoding process 404 encodes the auxiliary information into the content.  A transmitter 406 then transmits or broadcasts the watermarked content to a viewer.

20    The watermark encoder may encode auxiliary information throughout the entire video frame in which at least one marked video object resides.  For example, the user may specify via the editing tool the location of two or more video objects by drawing a boundary around the desired video objects in a video sequence.  The encoding process records the screen location information for each object in the relevant frames and

25    associates it with the auxiliary information provided by the user, such as an object identifier.  The encoder then creates a watermark message for each frame, including the screen location of an object for that frame and its object identifier.  Next, it encodes the watermark message repeatedly throughout the frame.

An alternative approach is to encode auxiliary information for an object in the screen location of each frame where a video object layer for that object resides (described fully in figure 6 below).

Fig. 5 is a diagram depicting yet another process for embedding auxiliary

5   information about video objects in a video stream. This process is similar to the one shown in Fig. 4, except that the position of video objects is derived from transmitters 500-504 attached to the real world objects depicted in the video scene and attached to video cameras. The transmitters emit a radio signal, including an object identifier. Radio receivers 506 at fixed positions capture the radio signal and provide information to a pre-

10   processor 508 that triangulates the position of each transmitter, including the one on the active camera, and calculates the screen location of each transmitter in the video stream captured by the active camera. The active camera refers to the camera that is currently generating the video stream 510 to be broadcast or transmitted live (or recorded for later distribution). In a typical application, there may be several cameras, yet only one is

15   selected to provide the video stream 510 at a given time.

Next, an encoding process 512 selects video objects for which auxiliary information is to be embedded in the video stream. The selection process may be fully or partially automated. In a fully automated implementation, a programmed computer selects objects whose screen location falls within a predetermined distance of the 2D

20   screen extents of a video frame, and whose location does not conflict with the location of other objects in the video frame. A conflict may be defined as one where two or more objects are within a predetermined distance of each other in screen space in a video frame. Conflicts are resolved by assigning a priority to each object identifier that controls which video object will be watermark enabled in the case of a screen location

25   conflict.

In a partially automated implementation, the user may select one or more video objects in frames of the video stream to be associated with embedded watermark information via a video editing system 514. The video editing system may be implemented in computer software that buffers video frame data and associated screen

location information, displays this information to the user, and enables the user to edit the screen location information associated with video objects and select video objects for watermark encoding.

After calculating video object locations and selecting them for watermark encoding, a watermark encoding process 516 proceeds to encode an object identifier for each selected object. The watermark may be encoded in screen locations and frames occupied by a corresponding video object. Alternatively, object identifiers and corresponding screen location information may be encoded throughout the video frames (or in the audio track of an audio visual work).

After watermark encoding, a transmitter 518 transmits or broadcasts the video stream to viewers. The video stream may also be stored, or compressed and stored for later distribution, transmission or broadcast. The watermarks carrying object identifiers, and other object information, such as screen location information, may be encoded in uncompressed video or audio, or in compressed video or audio.

Fig. 6 depicts an example watermark encoding process that may be used in some of the systems described in this document. Depending on the implementation, some of the processing is optional or performed at different times. The watermark encoding process operates on a video stream 600. In some cases the stream is compressed, segmented into video object layers, or both compressed and segmented into video objects as in some video content in MPEG 4 format. The encoder buffers frames of video, or segmented video objects (602).

In this particular example, the encoder embeds a different watermark payload into different portions of video frames corresponding to the screen location of the corresponding video objects. For example, in frames containing video object 1 and video object 2, the encoder embeds a watermark payload with an object identifier for object 1 in portions of the frames associated with object 1 and a watermark payload with object identifier for object 2 in portions of the frames associated with object 2. To simplify decoder design, the watermark protocol, including the size of the payload, control bits, error correction coding, and orientation/synchronization signal coding can be the same

throughout the frame. The only difference in the payloads in this case is the object specific data.

A variation of this method may be used to encode a single watermark payload, including identifiers and screen locations for each watermark enabled object, throughout each frame. While this approach increases the payload size, there is potentially more screen area available to embed the payload, at least in contrast to methods that embed different payloads in different portions of a frame.

Next, the encoder optionally segments selected video object instances from the frames in which the corresponding objects reside. An input to this process includes the screen locations 606 of the objects. As noted above, the screen locations may be provided by a user via a video editing tool, or may be calculated based on screen location coordinates derived from transmitters on real world objects. The screen extents may be in a coarse form, meaning that they do not provide a detailed, pixel by pixel definition of the location of a video object instance. The screen extents may be as coarse as a bounding rectangle or a polygonal shape entered by drawing a boundary around an object via a video editing tool.

Automated segmentation may be used to provide refined shape, such as binary mask. Several video object segmentation methods have been published, particularly in connection with object based video compression. The implementer may select a suitable method from among the literature that satisfies the demands of the application. Since the watermark encoding method may operate on blocks of pixels and does not need to be precise to the pixel level due to human interaction, the segmentation method need not generate a mask with stringent, pixel level accuracy.

In some implementations, video objects are provided in a segmented form. Some examples of these implementations are video captured of a physical object (e.g., actor, set, etc.) against a green screen, where the green color of the screen helps distinguish and define the object shape (e.g., a binary mask where a given green color at a spatial sample in a frame indicates no object, otherwise, the object is present).

Next, the encoder computes a bounding region for each object (608), if not already available. The bounding region of a video object instance refers to a bounding rectangle that encompasses the vertical and horizontal screen extents of the instance in a frame. The encoder expands the extents to an integer multiple of a watermark block size (610). The watermark block size refers to a two dimensional screen space in which the watermark corresponding to a video object, or set of objects, is embedded in a frame at a given encoding resolution.

The watermark encoder then proceeds to embed a watermark in non-transparent blocks of the bounding region. A non-transparent block is a block within the bounding region that is not overlapped by the video object instance corresponding to the region. The watermark for each block includes an object specific payload, such as an object identifier, as well as additional information for error correction and detection, and signal synchronization and orientation. The synchronization and orientation information can include message start and end codes in the watermark payload as well as a watermark orientation signal used to synchronize the detector and compensate for changes in scaling, translation, aspect ratio changes, and other geometric distortions.

There are many possible variations to this method. For example, an object specific watermark may be encoded throughout a bounding rectangle of the object. This approach simplifies encoding to some extent because it obviates the need for more complex segmentation and screen location calculations. However, it reduces the specificity with which the screen location of the watermark corresponds to the screen location of the video object that it is associated with. Another alternative that gives fine screen location detail, yet simplifies watermark encoding is to embed a single payload with object identifiers and detailed location information for each object. This payload may be embedded repeatedly in blocks that span the entire frame, or even in a separate audio track.

In some watermark encoding methods, the watermark signal may create visible artifacts if it remains the same through a sequence of frames. One way to combat this is to make the watermark signal vary from one frame to the next using a frame dependent

watermark key to generate the watermark signal for each block. Image adaptive gain control may also be used to reduce visibility.

## *Decoding Processes*

There are a variety of system configurations enabling users to access watermark
5   enabled features in video objects. Before giving some examples, we start by defining decoder processes. The examples then illustrate specific system configurations to implement these processes.

As depicted in Fig. 7, there are five principal decoding processes: 1) decoding auxiliary information embedded in a watermark in the video content (700, 702); 2) user
10  selection of watermark enabled information or actions (704); 3) determining information or actions associated with a video object (706); and 4) rendering watermarked enabled information or actions to the user (708). Rendering may include generating visual, audio or audio-visual output to present information and options for selecting more information or actions to the user, executing a program or machine function, or performing some
15  other action in response to the watermark data.

The first process extracts auxiliary information, such as object identifiers and screen locations, from the video stream or an accompanying audio track. The next process implements a user interface to indicate to the user that the video has watermark enabled objects and to process user input selecting watermark enabled information or
20  actions. The third process determines the information or action associated with a selected video object. Finally, the fourth renders watermarked enabled information or actions to the user.

Each of these decoding processes need not be implemented in all applications. A decoder may operate continuously or in response to a control signal to read auxiliary
25  information from a watermark, look up related information or actions, and display it to the user. Continuous decoding tends to be less efficient because it may require a watermark decoder to operate on each frame of video or continuously screen an audio track. A more efficient approach is to implement a watermark screen that invokes a watermark decoder only when watermark data is likely to be present. A control signal

sent in or with the video content can be used to invoke a watermark decoder. The control signal may be an in-band signal embedded in the video content, such as a video or audio watermark. For example, a watermark detector may look for the presence of a watermark, and when detected, initiate a process of decoding a watermark payload,

5   accessing information or actions linked via an object identifier in the payload, and displaying the linked information or actions to the user. The control signal may be one or more control bits in a watermark payload decoded from a watermark signal.

The control signal may also be an out-of-band signal, such as tag in a video file header, or a control signal conveyed in a sub-carrier of a broadcast signal.

10   The control signal can be used to reduce the overhead of watermark decoding operations to instances where watermarked enabled objects are present. The decoder need only attempt a complete decoding of a complete watermark payload when the control signal indicates that at least one video object (e.g., perhaps the entire frame) is watermark enabled.

15   The control signal may trigger the presentation of an icon or some other visual or audio indicator alerting the user that watermark enabled objects are present. For example, it may trigger the display of a small logo superimposed over the display of the video. The viewer may then select the icon to initiate watermark decoding. In response, the watermark decoder proceeds to detect watermarks in the video stream and decode

20   watermark payloads of detected watermarks. Additionally, when watermark payloads for one or more objects are detected, the user interface can present object specific indicators alerting the user about which objects are enabled. The user can then select an indicator to initiate the processes of determining related information or actions and presented the related information or actions to the user.

25   Another way to reduce watermark decoding overhead is to invoke watermark decoding on selected portions of the content in response to user selection. For example, the decoder may be invoked on portions of frames, a series of frames, or a portion of audio content in temporal or spatial proximity to user input. For example, the decoding process may focus a watermark decoding operation on a spatial region around a screen

location of a video display selected by the user. Alternatively, the user might issue a command to look for enabled content, and the decoding process would initiate a watermark detector on frames of video or audio content in temporal proximity to the time of the user's request. The decoding process may buffer frames of the most recently

5    received or played audio or video for the purpose of watermark screening in response to such requests.

### Example Configurations

One configuration is video player with an interactive user interface that displays video content and implements watermark enabled features. In this configuration, the

10    player decodes the watermark, displays video content, and enables the user to select video objects via its interactive user interface. The player may have a local database for looking up the related information or action of an identifier extracted from a video object.

Fig. 8 illustrates an example configuration of a decoding process for linking video objects to auxiliary information or actions. In this configuration, there are three primary

15    systems involved in the decoding process: 1) A local processing system (e.g., PC, set-top box, stand-alone device) 800 responsible for receiving video content, playing it on a display, and decoding watermarks from the content. 2) A router 802 that communicates with the local processing system via a network 803 such as the Internet; and 3) a web server 804 that communicates with the local processing system and the router via the

20    network. The local processing system may be implemented in a variety of consumer electronic devices such as a personal computer (PC), set-top box, wireless telephone handset, television, etc. The router and web server may similarly be implemented in a variety of systems. In typical Internet applications, the router and web server are implemented in server computers. For these applications, communication of data among

25    the local processing system, router and server may be performed using network protocols, such as TCP/IP, and other application level protocols such as XML, HTTP, and HTML.

The local processing system 800 receives a video stream 806 via a receiver 808. The type of receiver depends on the nature of the video transmission, such as Internet download or streaming delivery, satellite broadcast, cable television broadcast, television

broadcast, playback from portable storage device such as VHS tape, DVD, etc. In each

case, an appropriate device, such as network adapter, satellite dish, tuner, DVD driver,

etc. receives the content and converts it to a video signal. This process may also included

decompressing a compressed video file. However, as noted above, the watermark may be

5    encoded and decoded from compressed video or audio, such as MPEG 4 video objects or

audio.

The local processing system renders the video content 810. In a PC, the rendering

process includes converting the video signal to a format compatible with the video

controller in the computer and writing the video to video memory in the video controller

10    812. The video controller 812 then displays the video signal on a display device 814.

As the video is being rendered, the local processing system buffers frames (816)

of audio or video for watermark detecting and decoding. In a PC, the buffering may be

integrated with rendering the video to video memory or may be implemented as a

separate process (e.g., allocating separate video buffers in main memory or video

15    memory). Also, depending on the nature of the video signal and encoding process, the

buffer may store frames of compressed video content or decompressed video content

from which watermarks are detected and decoded.

A watermark detector screens the buffered content for the presence of a

watermark (818). If a watermark is present, it sends a message to a user interface

20    application 820, which in turn, generates a graphical logo or other visual or audio signal

that indicates the presence of watermarked enabled video objects.

A watermark decoder 822 reads one or more watermark payloads from the

content. As noted above, the decoder may be triggered by one or more of the following

events: 1) the detector finding the presence of a watermark; 2) an out-of-band control

25    signal instructing the decoder to detect and decode a watermark; 3) user selection of the

graphical logo, etc.

In addition to displaying an indicator of watermark enabled objects, the user

interface 820 also manages input from the user for selecting video objects and for

controlling the display of information associated with selected video objects. In a PC

environment, the user interface can be implemented as an interactive display with graphics that respond to input from a gestural input device, such as a mouse or other cursor control device, touch screen, etc. This interactive display is superimposed on the display of the video stream. In this environment, the user selects a video object by

5    placing a cursor over the video object on the display and entering input, such as clicking on a mouse.

The specific response to this input depends on the implementation of the watermark decoder and how the content has been watermarked. In one class of implementations, the watermark payload contains information for each watermark

10    enabled object in the video content, along with a location codes specifying screen locations of the objects. In this type of implementation, the decoder preferably decodes the watermark payload in response to detecting presence of a watermark and stores the payload for the most recently displayed video content. In response to user input selecting a video object, the decoder receives the coordinates of the user selection and finds the

15    corresponding location code in the watermark payload information that defines a screen area including those coordinates. The location code is specified at a reference frame resolution, and the user selection coordinates are normalized to this reference resolution.

In another class of implementations, video frames contain one or more watermarks, the payloads in those watermarks are specific to the video objects in which

20    they are embedded.

There are a couple of alternative ways of mapping the location of a user selection to a corresponding watermark payload. One approach to decoding the video frame is to decode watermark payloads for each watermark detected in the frame, and then store screen location data indicating the location of the watermark containing that payload.

25    The screen coordinates of a user's selection can then be mapped to a payload, and specifically to the object identifier in the payload, based on the screen location data of the watermark.

Another approach to decoding is to execute a decode operation on a specific temporal and spatial region in proximity to the temporal and spatial coordinates of a user

selection. The temporal coordinates correspond to a frame or set of frames, while the spatial coordinates correspond to a two-dimensional region in the frame of set of frames. If the decoder can decode a watermark payload from the region, then it proceeds to extract the object identifier and possibly other information from the payload. If the

5      decoder is unsuccessful in decoding a payload from the region, it may signal the user interface, which in turn, provides visual feedback to the user that the attempt to access a watermark enabled feature has failed, or it may search frames more distant in time from the user's selection for a watermark before notifying the user of a failure.

The watermark decoder can enhance the user's chances of selecting a

10     watermarked enabled object by providing graphical feedback in response to user selection of the video frame or object within the frame. For example, the decoder can give the user interface the screen coordinates of areas where a watermark has been detected. Screen areas that correspond to different watermark payloads or different object locations as specified within a watermark payload can be highlighted in different

15     color or some other graphical indicator that distinguishes watermark enabled objects from unmarked objects and each other.

The decoder forwards an object identifier (824) for the video object at the selected location to the server 802 via a network interface 826. The decoder may also provide additional information from the watermark or context information from the local

20     processing system. For Internet applications, the decoder sends a message including this information to the server in XML format using HTTP. Before forwarding the message, the user interface may be designed to prompt the user with a dialog box requesting the user to confirm that he or she does want additional information.

The network interface 826 forwards the message to the server 802 over the

25     network. While this example is particularly directed to computer networks like the Internet, similar systems may be built for other types of networks, like satellite broadcast networks, wireless phone networks, etc. In these types of networks, the network interface corresponds to the device and accompanying programming that sends and receives data over a communication link. In the case of wireless device, the network interface may be

a cellular telephone transceiver. In the case of the satellite broadcast network, the network interface may be a satellite dish. Note that combinations of technologies may be used for transmitting and receiving functions, such as sending data via telephone network using a modem or network adapter, and receiving data via a satellite dish.

5       The server, in response to receiving the message (828), parses it and extracts an index used to look up a corresponding action in a database (830) that associates many such indices to corresponding actions. The index may include the object identifier and possibly other information, such as time or date, a frame identifier of the selected object, its screen location, user information (geographic location, type of device, and

10    demographic information), etc. Several different actions may be assigned to an index. Different actions can be mapped to an object identifier based on context information, such as the time, date, location, user, etc. This enables the server to provide actions that change with changing circumstances of the viewer, content provider, advertiser, etc. Some examples include returning information and hyperlinks to the user interface 820

15    (e.g., a web page), forming and forwarding a message to another server (e.g., re-directing an HTTP request to a web server), recording a transaction event with information about the selected object and user in a transaction log, downloading to the local processing system other media such as still image, video or audio content for playback, etc.

      Another action that may be linked to the video object is connecting the user to a

20    transaction server. The transaction server may enable the user to purchase a physical object depicted in the video object via an electronic transaction. It may also enable the user to enter into a contract electronically to obtain usage rights in the video content or related content.

      In the example configuration depicted in Fig. 8, the server 802 looks up the

25    address of a web server associated with the index (830). It then forwards an HTTP request (832) to the web server 804 at this address and provides the IP address of the local processing system 800. In addition, it may also include in the HTTP request that the web server may use to tailor a response to the local processing system, such as the object identifier, frame identifier, user demographics, etc.

The web server receives the request (834) and returns information to the local processing system (836). This information may include hyperlinks to other information and actions, programs that execute on the local processing system, multimedia content (e.g., music, video, graphics, images), etc. One way to deliver the information is in the

5     form of an HTML document, but other formats may be used as well.

The local processing system receives the information from the server 804 through the network and the network interface 826. The decoder operates in conjunction with the user interface application such that the information is addressed to the user interface. For Internet applications, a TCP/IP connection is established between the user interface

10    application and the network. The server forwards the information to the IP address of the user interface application. The user interface then formats the information for display and superimposes it onto the video display. For example, when the information is returned in the form of HTML, the user interface application parses the HTML and formats it for display on display device 814. The rendered HTML is layered onto the

15    video frames in the video memory. The video controller 812 then displays a composite of the HTML and the video data. In the event that the HTML includes hyperlinks, the user interface processes inputs to these links in a similar fashion as an Internet browser program.

Just like the servers may map a watermark payload to different actions for

20    different circumstances, the user interface may also implement a set of rules that govern how it presents content returned from the network based on context information. For example, the user interface may keep track of information that a user has scene before and change it or tailor it based on user information or user preferences entered by the user. For example, the user can configure the user interface to display information about

25    certain topics (news categories like sports, business, world affairs, local affairs, entertainment, etc.) or actions (e.g., links to certain categories of electronic buying transactions, video or music downloads, etc.). Then, when the user interface receives information and links to actions, it filter the information and links based on user preference and provide only information and links in the user's preference.

One potential drawback of the above configuration is that it may create conflicts among viewers. People often watch TV in a shared environment, whereas they work on the Internet in a personal environment. This environment creates a conflict when one viewer selects an object to get information that interferes with another viewer's

5    enjoyment of the video program.

One solution is to provide consumers with their own personal and portable Internet personal device (PD) as shown in Fig. 9. The system may be configured to have the decoding process in a TV, set-top box, or other receiver 900 of a video stream. The decoder may then transmit watermark IDs, locations, and potentially other context

10   information to the PD 902.

As another alternative, the decoder may be located in the PD. For example, the PD may be equipped with a microphone that captures the audio signal emitted from the speaker of the television. The PD digitizes the audio signal and extracts watermarks from it, which include object information used to link video objects to information or actions.

15   For example, the object information may include object identifiers and location codes for video objects in the video program. The PD may also include a camera, and perform similar actions on watermarks in the video frames.

Two parts of this configuration are: 1) a transmitting device like the television 900 shown in Fig. 9, set-top box, etc., and 2) a receiving PD 902 such as a personal

20   digital assistant (PDA) with a wireless connection to the Internet, or a remote control. The receiving PD can perform the functions of enabling the user to select a video object, retrieving the linked information or actions for the selected object, and rendering them on its user interface. One example of such a device is a PD with a communication link (e.g., infrared, radio, etc.) to the transmitting device for receiving object information and a

25   communication link with a network, database, server, etc. for retrieving the linked information or actions for the selected object. As another alternative, the receiving PD acts solely as a user control device of the transmitting device that enables the user to select an object and communicates the selection back to the transmitting device. The transmitting device, in response to the user selection, retrieves linked information or

actions for the selected object and renders them. One example of such a device is a remote control with a user interface (e.g., display and cursor control device for selecting objects) and a two-way communication link with the transmitting device (e.g., infrared, radio, etc.).

5        ### *Transmitting Device*

The transmitter could be a stand-alone device or part of a set-top box that already exists for your TV. The stand-alone device can be a small transmitter that attaches to coaxial cable and transmits a video object identifier and its location during the TV show. If this stand-alone device is connected before the channel has been chosen, it can transmit

10       the IDs and locations for all channels, and the receiving PD can be used to choose the channel you are watching. Alternatively, the receiving PD can transmit an identifier of the channel you are watching to the transmitting device, so it, in turn, only transmits the information for the desired channel.

A less complex stand-alone solution, thus less expensive to manufacture and sell,

15       is to add this stand-alone device after the channel has been chosen, possibly between your VCR or set-top box and your TV, and have it transmit information for the channel you are watching. Finally, this stand-alone device can be OEM hardware that is added inside the TV by the manufacturer or as a post buying solution (i.e. retro-fit).

The set-top box solution may use a Web, Cable or Digital TV set-top box,

20       especially if the existing box is already interactive. Otherwise, OEM hardware could be provided for the set-top box manufacturer.

The transmission scheme can use any method, such as IR or radio waves (e.g., Bluetooth wireless communication), to transmit this minimal amount of information. IR ports are advantageous because most laptops and PDAs already have IR ports. If the set-

25       top box already has a transmission protocol, the transmission scheme should use that scheme. If this scheme is not applicable with an existing receiving PD, a special attachment can be developed and feed into the receiving PD via existing input devices, such as IR, serial, parallel, USB, or IEEE firewire inputs.

### Receiving PD

The receiving PD may be a laptop computer, Palm pilot, digital cell phone, or an Internet appliance (such as a combined PDA/Cell Phone/Audio/Video device). This PD would display the links in their relative location on a screen matching the TV screens

5   aspect ratio. Then, using the PD you can select the desired link, possibly by clicking on the link, pressing the appropriate number key relating to the link number, or saying the link number and using speech recognition (906). Next, the PD sends information about the selected link to a database (e.g., a web server that converts the information into a web page URL and directs the server at this URL to return the corresponding web page to the

10  PD) (908). A user interface application running in the PC then renders the web page (910) on its display. Using this approach, the links are dynamic and the data required to describe a link is minimal. This allows the watermarking and transmitting process to be easier. Most importantly, fewer bits need to be transmitted since only an ID and not the complete link are required.

15  Alternatively, if the receiving PD is connected to the Internet, new and hot information can automatically be pushed to the receiving PD, rather than requiring the user to click on the link. For example, if you are watching a basketball game, the current stats of the player with the ball can be pushed. Or, if you are watching a concert, the location on the tour can be presented. This push feature can be always-on or controlled

20  by the user.

The configuration shown in Fig. 9 differs from the one shown in Fig. 8 in that decoding of a watermark payload and user selection of a link associated with that payload are performed on separate devices. The functions of receiving and rendering video content, decoding watermark from the content, and linking to information and actions

25  based on the watermark payload can be performed on separate devices. Many of the features and applications detailed in connection with Fig. 8 also apply to the configuration shown in Fig. 9.

The following sections illustrate several different application scenarios and related watermarking systems and methods that demonstrate the diversity of the technology described above.

### Previously Segmented Video

5   Segmented video streams, such as those supported in MPEG 4 allow the film or video editor to extract a video scene element from the background and embed the isolated video object. The watermark encoder marks a video object layer corresponding to the object in some or all frames in which the object is visible. When the scene element is not large enough to be encoded with at least one watermark block, the editor keys in that

10  frame, defines a new element again and begins a batch embedding along each frame of the time sequence.

The viewer will watch the movie on DVD, VHS, or some other video signal format and be able to link directly to the Internet or other database online or offline by selecting a watermark enabled video object.

### Video Objects Captured Through Greenscreeens

15  The embedding process may embed a live character that has been shot against a greenscreen. This enables a video editor to embed the actor without first extracting him from the background. This video object will later be composited with computer graphics or other live action shot at another time. Watermark embedding technology described

20  above can be integrated with commercially available video compositing software from Discreet Logic, Adobe or Puffin Designs.

### Rendered 3D Object Layers

Watermarks may also be embedded in two dimensional image renderings of still or animated 3D graphical objects. The embedded object can be composited with a video

25  stream to form a video program, such as a movie or television programming. This embedded object stays in the video content when converted to other formats such as DVD or VHS without an additional watermark embedding. Conversely, graphical

objects that link to information or electronic commerce transactions can be added to a
video product, such as a movie, when its converted from one format to another. For
example, the video content can be watermark enabled when it is placed on a DVD or
VHS for mass distribution.

5

### Physical Objects Captured in Video

Another application is to embed video objects that are static like the basketball
backboard or the sportscaster's table or the Jumbotron. This entails masking out the
static video object layer in each frame to isolate it from the background in the video
10    sequence. This may be accomplished by creating two separate video feeds from the same
camera using one to create the mask for each "frame" and using the other for the actual
broadcast signal. The masked area is marked and the two signals are combined and
broadcast.

The sportscaster's table could also have a watermark on the actual artwork that
15    scrolls in front of it. This persistent watermark would need no additional masking.

### Real Time Object Embedding

Another application is to embed video objects such as the players of a game.
Using video object segmentation, this application extracts video objects from the
background and embeds them in the video stream before broadcast or other distribution.
20    Another method is to generate different video streams, each potentially including
a different watermark or watermark payload linking video objects in the corresponding
video stream to actions or information. In this case, a watermark is embedded in the
video captured from a camera that focuses on a particular character, player, or object. In
a video production process, a technician selects the video feed from this camera from
25    among feeds from one or more other cameras to be part of the final video program. For
example, a camera following a particular player is encoded with an object identifier
associated with that player. The technician selects the video feed from this camera (e.g.,
the Kobe Kamera isolated on the Laker's Kobe Bryant) at intervals during a game and

carries the watermark enabling the user to click the frame and access a page of a web site like NBA.com, Lakers.com, etc. that provides information about that player.

Also, a transparent frame could be overlaid on this camera that the view could not see, but the detector could. Just enough pixels would be sent to detect the image.

5       Yet another method is to compute video objects dynamically at video capture by deriving video object position and screen extents (bounding box, binary mask, shape, etc.) from the real world objects being captured.

### Games

Watermarks may be inserted into graphical objects in 3D animation used in video 10 games to link characters and other objects to information or actions. Dreamcast, Playstation 2, and PC CD-ROM games all have Internet access. Images that are rendered on the fly can be embedded with the watermark. Canned animation and cut scenes are rendered previously with the watermark in them. These can activate special website interaction, or for playing online, this could allow extra interaction between players.

### Embedding Graphic Overlays

The score area on the bottom of the screen is an excellent place to mark before transmission of the video broadcast.

Real Time embedding is ready for delivery. Every NFL and NBA broadcast now has sophisticated graphics that are keyed on screen.

20       In addition, another opportunity to mark is when a player's statistics are shown on the NFL game between plays or during a timeout. The screen cuts from the live broadcast to canned animation that includes a composite of the player's picture and his states. This is an excellent opportunity for watermark embedding.

In addition to the real time embedding examples above, one method is to embed a 25 watermark or watermarks in relatively static portions of the background (e.g., watermarking portions of video frames depicting the turf of a playing field). This method would work well since it is stationary and usually fills a large part of the TV screen.

### News Broadcasts

Graphics used in news broadcasts can be linked to information and actions via watermarks. CNN, ABC, NBC, CBS, etc. have used keyed images over the anchor's shoulder for years. They are canned graphics that are composited during the broadcast. These canned graphics can be embedded with watermarks as described above.

### Virtual Billboards

The virtual billboards displayed advertising from the typical broadcast advertiser. These images can be watermarked to link the virtual billboards to information or actions, like electronic buying opportunities.

### Feature Films

Feature films that were not embedded in the original post-production can be embedded afterwards on their way to video, DVD, or other format for electronic or packaged media distribution.

### Logos and other Graphic Overlays

Many channels now keep a logo at the bottom right corner of their screen. The History Channel, MTV, VH1, TLC, TNN, all have logos that advertise the channel. These logos are sometimes shown throughout the program hour. These logos can be linked to external actions or information by embedding a watermark in either the video signal or the accompanying audio track.

### Watermarked Signs

Watermarks may be embedded in the images on large physical objects, such as outdoor signs. These outdoor signs could conceivably be marked and detected onscreen. A typical example would be billboards inside a baseball park or football stadium. When video is captured of these physical objects, the watermarked images on these objects is recorded in the video signal. The watermark is later decoded from the video signal and used to link the video signal to an action or information.

### Watermark Enabled Advertising

Video objects representing advertising or promotions may be watermark enabled. For example, an advertiser such as Ford would produce a watermark enabled ad that would pop up specifically for users to click. The promo could be "NFL on

5    ESPN...Brought to You By FORD" and while that logo or graphic spins there for twenty seconds Ford is offering a promotional discount or freebie for all the people that click on it to visit there site during that time. The video programmer could run the video objects many times so people who miss it could get another chance.

### User Alerts and Preferences

10    The watermark decoding system may employ a user interface to enable the user to control activation of watermark enabled features. For example, the decoding process may default to an "alert off" status, where the watermark decoder does not alert the user to watermark enabled features unless he or she turns it on. By querying the screen every few seconds, a watermark detector or decoder may alert the user that there are watermark

15    enabled objects present on screen if he/she so chooses. The decoding system may be programmed to allow the user to determine whether or not he/she is alerted to watermarked enabled features, and how often.

In addition, the decoding system may enable the user to set preferences for certain types of information, like sports, news, weather, advertisements, promotions, electronic

20    transactions. The decoding system then sets up a filter based on preferences entered by the user, and only alert the user to watermark enabled features when those features relate to the user's preferences.

### Watermark Enabled Commerce

Watermark enabled video objects may be linked to electronic commerce and

25    advertising available on the Internet or from some other information server.

For example, video objects may be linked to opportunities to rent or by the content currently being viewed or related content. At the beginning or end of the film, a watermark enabled logo may be overlayed on a video signal (e.g., from a DVD or other

video source) to allow the user to access a website to review the movie, purchase the movie (rent to own), rent/buy the sequel, alert the web site that the rented movie has been viewed to help manage inventory, etc.

### Introducing Interactivity into Video Programming

5          By incorporating watermark enabled video into a television program, the program may be transformed into an interactive experience. For example, a sitcom program could include watermark enabled video objects at selected points in the broadcast or at the opener that alerted the viewer to get online.

### Interactive Shopping

10          Video advertising of products, such as clothing, may be watermark enabled to link video objects representing a product or service to additional information or actions, such as electronic buying transactions. For example, a clothing manufacturer could enable all their broadcast ads. Each piece of clothing on the actor may be watermark enabled and linked to the page on the web site to buy the article.

### Real Time Derivation of Video Object Spatial and Temporal Extents

15          The technology shown in Fig. 5 allows watermark tracking by placing locator devices in physical objects. One example is to place these locators inside the shoes and on the uniforms of professional athletes during games. These locator chips emit a signal that is received and triangulated by detectors on courtside. Each chip has a unique ID to

20     the player. The signal is passed through a computer system integrated into the production room switcher that embeds watermarks into the video stream captured of the player.

          The players wear at least two transmitters to give location information relative to the camera position. Using this information, a preprocessor derives the screen location of the corresponding video objects. If transmitters get too close to distinguish a video

25     object, the preprocessor prioritizes each video object based on the producer's prior decision.

Alternatively, the player's jersey could be watermarked, and used like a pre-marked static object.

### Linking Audio Objects with Watermarks

Just as audio or video watermarks can be used to link video objects to information

5    or actions, so can they link audio objects to related information or actions.  In an audio signal, portions of the signal are distinguishable and recognizable as representing a particular audio source, such as a person's voice or vocal component of a song, an instrument, an artist, composer, songwriter, etc.  Each of these distinguishable components represent audio objects.  Watermarks in the audio or accompanying video

10   track can be used to link audio objects to information or actions pertaining to the action.

To access linked information or actions, the user selects a portion of the audio signal that includes a watermark enabled audio object, such as by pressing a button when an audio object of interest is currently playing.  Using the temporal location of the user selection in the audio signal, a watermark linking process maps the user selection to a

15   corresponding audio object.  The systems and processes described above may be used to retrieve and render information or actions linked to the selected audio object.

### Concluding Remarks

Having described and illustrated the principles of the technology with reference to

20   specific implementations, it will be recognized that the technology can be implemented in many other, different, forms.  To provide a comprehensive disclosure without unduly lengthening the specification, applicants incorporate by reference the patents and patent applications referenced above.  These patents and patent applications provide additional implementation details.  They describe ways to implement processes and components of

25   the systems described above.  Processes and components described in these applications may be used in various combinations, and in some cases, interchangeably with processes and components described above.

The methods, processes, and systems described above may be implemented in hardware, software or a combination of hardware and software. For example, the watermark encoding processes may be incorporated into a watermark or media signal encoding system (e.g., video or audio compression codec) implemented in a computer or computer network. Similarly, watermark decoding, including watermark detecting and reading a watermark payload, may be implemented in software, firmware, hardware, or combinations of software, firmware and hardware. The methods and processes described above may be implemented in programs executed from a system's memory (a computer readable medium, such as an electronic, optical or magnetic storage device). Additionally, watermark enabled content encoded with watermarks as described above may be distributed on packaged media, such as optical disks, flash memory cards, magnetic storage devices, or distributed in an electronic file format. In both cases, the watermark enabled content may be read and the watermarks embedded in the content decoded from machine readable media, including electronic, optical, and magnetic storage media.

The particular combinations of elements and features in the above-detailed embodiments are exemplary only; the interchanging and substitution of these teachings with other teachings in this and the incorporated-by-reference patents/applications are also contemplated.